OE FORM 6000, 2/69

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

ERIC ACC. NO.				
ED 039 395				
CH ACC. NO.	P.A.	PUBL. DATE	ISSUE	
AA 000 573		67	RIEOCT70	

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒
ERIC REPRODUCTION RELEASE? YES ☐ NO ☒
LEVEL OF AVAILABILITY I ☒ II ☐ III ☐

AUTHOR
Pasanella, Ann K.; And Others

TITLE
Bibliography of Test Criticism.

SOURCE CODE	INSTITUTION (SOURCE)
QPX15900	College Entrance Examination Board, New York, N.Y.
SP. AG. CODE	SPONSORING AGENCY

EDRS PRICE	CONTRACT NO.	GRANT NO.
0.25; 2.90		
REPORT NO.		BUREAU NO.

AVAILABILITY

JOURNAL CITATION

DESCRIPTIVE NOTE
56p.

DESCRIPTORS *Annotated Bibliographies; *Evaluation Criteria; *Educational Testing; *Test Validity; *Test Reliability; Test Selection; Test Construction; Testing; Conformity; Test Results; Academic Freedom; Test Interpretation; Civil Liberties; Testing Problems; Examiners; Measurement Instruments

IDENTIFIERS College Entrance Examination Board

ABSTRACT This is a selected compilation of 47 items relating to criticisms of tests and testing. The items cover the period of ten years immediately preceding the year 1966 and are held to be scholarly writings almost without exception. The books and articles listed carry extensive annotations and focus on the following aspects of tests and testing: encouragement of intellectual conformity; erosion of individual freedom of choice; exertion of undue influence on education; invasion of individual privacy; and, concealment of true character by masquerading as scientific instruments. The sources of strain in the themes of these books and articles is thought to be three-sided: the tests themselves; the test users; and, the test makers. (RJ)

GPO 870-390

ED 039 395

Bibliography of Test Criticism

Ann K. Pasanella, Winton H. Manning, and Nurhan Findikyan
College Entrance Examination Board

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

AA 000 573

College Entrance Examination Board, 1967.

Bibliography of Test Criticism

Ann K. Pasanella, Winton H. Manning, and Nurhan Findikyan

"The central issues are the corrupting effects of multiple-choice tests on education, the manner in which the tests favor brilliant superficiality over depth, subtlety, and creativity, and the manner in which the very nature of the tests allows control of testing to fall into the hands of people whose approach to the admittedly formidable problem of testing is not so much that of the scholar as that of the cost accountant and the statistical technician."
(Banesh Hoffmann, in a letter to Science, March 6, 1964)

Although testing was born at approximately the same time as our century, it took over 50 years before an identifiable legion of test critics developed. Why the protest accelerated so slowly, who the critics are, and what shape the revolt will take would be interesting subjects for further study.

We have complied a selected annotated bibliography of the literature of test criticism over the last 10 years. We have concentrated on scholarly writings, and articles and features in popular magazines have, for the most part, been excluded. The reader will rapidly discover that these 47 articles and books have certain common themes: tests encourage intellectual conformity; tests erode individual freedom of choice; tests exert undue influence on education; tests invade individual privacy; tests masquerade as scientific instruments. The sources of strain are thought to be three-sided: tests themselves; test users; and test makers. Indeed, it seems that for many persons, the critics are so persuasive they are hard to resist.

It is our hope that the bibliography will provide useful background for the Commission on Test's June meeting. The voices of protest are not silent this spring.

May 19, 1967

American Psychological Association. "Ethical Standards of Psychologists," American Psychologist, Vol. 18, January 1963, pp. 56-60.

These ethical standards, expressed in 19 principles, are designed to promote the science of psychology while protecting the welfare of others.

The psychologist is committed to objectivity and integrity. He maintains high standards of professional competence, shows sensible regard for moral and legal standards, avoids misrepresentation of his own qualifications and purposes, and gives information with modesty, scientific caution, and due regard for the limits of present knowledge. The psychologist does not communicate information about an individual to others unless certain important conditions are met: express permission is given by the individual or there is clear and imminent danger to an individual if it is not revealed; evaluative data about children, students, employees are discussed only for professional purposes and only with persons clearly concerned with the case; confidentiality of records is ensured.

The psychologist protects the client's welfare by putting the client's interest first. The psychologist who requests personal information in the case of interviewing or testing does so "only after making certain that the responsi-

ble person is fully aware of the purposes of the interview, testing or evaluation and of the ways in which the information may be used." The psychologist protects the security of psychological tests and other assessment devices whose value depends in part on the naivete of the subject by restricting access to persons with professional interests who will safeguard their use. Test scores, like test materials, are released only to persons who are qualified to interpret and use them properly.

When test results are communicated to parents and students they are to be accompanied by adequate interpretive aids or advice. Test results used for evaluation or classification are communicated to appropriate persons in such a way as to guard against misuse. "In the usual case, an interpretation of the test result rather than the score is communicated."

Psychological tests are published in a professional way with suitable manuals. Among other points, these must contain descriptions of appropriate populations, qualifications required for test interpretation, and warnings about possible interpretations not yet substantiated by research.

The psychologist seriously considers the possibility of emotional harm from his research and conducts it only when the subjects are aware of the possibility and consent to participate nonetheless.

Anastasi, Anne, "Psychology, Psychologists, and Psychological Testing," American Psychologist, Vol. 22, April 1967, pp. 297-306.

Psychological testing is becoming dissociated from the mainstream of contemporary psychology. Refinements in test construction preoccupy psychometricians, who have lost sight of the behavior they set out to measure. This condi-

tion has been a principal reason for the prevalent hostility of the public toward testing. The antitest revolt is characterized by seven concerns: (1) psychological tests may represent an invasion of privacy; (2) communication of test results often betrays confidentiality of the results; test results are too often inadequately interpreted to the examinees, thus leading to harmful misconceptions; some evidence exists that tests may be sulf-fulfilling prophecies; (3) criticism of individual items and of test content has often been unrealistic, but some sophisticated criticisms of item forms have been overlooked; (4) although tests are often blamed for reflecting objectionable features of the criteria they are designed to predict, it is possible that tests have not kept pace with changes in these criteria over time; (5) questions of fairness of tests to culturally disadvantaged groups have generally not been well defined; the use of moderator variables should be vigorously pursued; (6) tests are believed to foster rigid, inflexible, permanent classifications of persons; (7) tests tend to perpetuate a narrow definition of ability. Recent developments within psychological theory are examined from the standpoint of their implications for these concerns and for the development of testing generally.

Angoff, William, "The College Board and the Superior Student," The Superior Student, Vol. 7, April 1965.

Three questions raised about the predictive adequacy of the SAT are answered. These are: Does the SAT, which is geared to a diverse and heterogeneous population, discriminate sufficiently at the very high and very low ends of the ability range? Does the SAT discriminate against the superior student who can perceive the inadequacy of the

answer intended to be the correct one and either prefers an incorrect alternative or omits the item? Does the SAT fail to identify the creative person or divergent thinker?

Evidence indicates that tests specifically designed for a narrow ability range result in improved reliability and validity, hence provide better discrimination in the upper ability range than the broad-range SAT. However, the gain in validity is so small that technical problems of calibration, administration, and routing do not make the use of such narrow-range tests worthwhile.

Evidence also indicates that superior students are not put at a disadvantage because of their alleged tendency to "see beyond an item." The relationship between SAT scores and grades is linear throughout the ability range.

Finally, there seems to be no evidence, pro or con, for clarifying the relationship of the SAT to creativity, as the concept of creativity is hard to define indeed. Data from one study indicate, however, that the SAT has contributed to the selection of students who are not only academically superior but outstanding in their extracurricular activities as well.

Ballinger, Stanley E., "Of Testing and Its Tyranny," Phi Delta Kappan, January 1963, pp. 176-180.

Ballinger, associate professor of education at Indiana University, summarizes Hoffmann's criticisms of multiple-choice testing. Ballinger agrees with many of these themes and stresses the danger of equating quantitative treatment with objectivity. Ballinger thinks that Hoffmann places too large a share of the blame on the test makers. If standardized tests are too routine, it may well be that teaching and testing in college courses is information-oriented.

If critical thinking were at the heart of most teaching today, how much room would there be for tests composed of simple recognition-type multiple-choice items? This does not free the test makers from responsibility but spreads the responsibility more widely.

In our society, tests are increasingly important as a means for identifying talent. How adequate are they as a means for achieving the desired ends? Tests seem, by their very nature, to be a conservative force for preserving the current system. Is this a desirable consequence? How does one find the basis for determining the presence of talent that has not yet had the opportunity to be developed? The talent of culturally deprived Americans, for example, is not going to be discovered by pencil-and-paper tests. We must reconstruct our social arrangements in order to do this.

Hoffmann's book, The Tyranny of Testing, is likely to play a very useful role in the public scrutiny of tests and policies controlling their use. Ballinger seconds the proposal that a commission of inquiry into current testing practices be established.

Barclay, James R., "The Attack on Testing and Counseling: An Examination and Reappraisal," Personnel and Guidance Journal, Vol. XLIII, September 1964, pp. 6-16.

This examination of the nature of criticisms of counseling and testing procedures discusses the following charges by critics: that counseling practice and the use of testing is a Communist-inspired plot to subvert and pervert the morals of American youth; that testing is being misused by many so-called professionals and some individuals who are far from being professional; that some tests are personally

obnoxious to certain segments of the population and contain items that actually inform children of antisocial or law-breaking conduct; that the prediction from some of these tests is nearly null for individuals; and that there has been a widespread invasion of personal rights through the use of certain types of tests and the dissemination of these test results.

Barclay thinks some of these charges do represent deficiencies in current professional conduct and in training programs, although the critics often show personal bias and use faulty logic. Membership in professional organizations, a clearer understanding of the use of testing, some new considerations in counselor training, and a systematic program to inform the public are suggestions for answering the critics and improving both the practice of counseling and the use of testing procedures.

Barr, Donald, "A Note on the Technology of Cynicism," Columbia University Forum, Vol. VI, Summer 1963, pp. 32-38.

"Paired with almost every human task -- every enterprise that calls on us to accept responsibility. . . to discipline ourselves. . . to take guilt as well as glory to ourselves -- there is a machine task, similar in general appearance, but offering us the moral prophylaxis of prepared routine." When we must choose between the "human" and the "mechanical" performance of a task, we persistently choose the latter over the former. This is most pathetically illustrated in the multiple-choice test, which mechanizes "the most beautiful and subtly bold of all human enterprises, the education of the young."

The multiple-choice test dominates American education,

culminating in a panic over admission to college, in the center of which massively sits the College Entrance Examination Board. The Board's tests unfairly favor those who have learned the multiple-choice technique. Fortunately, or unfortunately, some fail to learn because they are not bright enough or because the technique conflicts with their ethical training.

The massive organizations behind testing are powerful in their protection of their methods and powerless to get beyond the "metaphysics of their mode of inquiry." A commission of inquiry is the only instrument possible to break out of the cynicism that pervades the testing enterprise.

Barzun, Jacques, The House of Intellect. New York: Harper & Brothers, 1959, 276 pp.

There are three enemies of the house of intellect: art, which claims of its devotees exclusive allegiance; science, which reserves the right to apply its method where it chooses; and philanthropy, which leaves no one alone.

Educators lay claim to the results of a science called educational research, but, in fact, no such science exists. "Human capacity is more varied than educational researchers know, but their methods insure that they shall never find this out." These researchers count events and score test papers, then derive meaningless generalities that extinguish any sparks of intellect in the classroom.

With mass education, the so-called "technique of educational measurement" is spreading. It attempts in an unsuccessful way to ape the language and methods of physical science. But, whether or not these educational tests can be considered scientific, the inexactitude of science when it deals with individuals is a subject that deserves the

attention of all who understand the obligation of intellectual rigor.

Visual memory is not the same as the power to summon up ideas. The power to summon up images by means of words is woefully neglected in our schools. Taking an objective test is simply pointing to ideas. It calls for the least effort of mind possible, that of recognition. There is no surprise, no fresh unfolding, but only the routine sorting out of the absurd and the trivial. "No other single practice explains more fully the intellectual defects of our students up to and through graduate school than their ingrained association of knowledge and thought with the scratching down of check marks on dotted lines."

A special appendix, prepared by Banesh Hoffmann analyzes some of the imprecisions and inconsistencies in multiple-choice tests, as illustrated in the Board's descriptive booklet for the Scholastic Aptitude Test.

Baumrind, Diana, "Some Thoughts on Ethics of Research: After Reading Milgram's 'Behavioral Study of Obedience,'" American Psychologist, Vol. 19, June 1964, pp. 421-423.

An experimenter has an ethical responsibility to his subjects. This is particularly crucial in a situation in which the experimental conditions expose the subject to loss of dignity or offer him nothing of value. The subject who volunteers for an experiment agrees implicitly to assume a posture of trust and obedience to the experimenter. But he has the right to assume that his security and self-esteem will be protected. This is not always done. The psychologist is only justified in exposing human subjects to emotional stress or other possible harm when the research problem is significant and can be investigated in no other way. Where

there is the danger of serious aftereffect, including loss of dignity, self-esteem, and trust in rational authority, research should be conducted only when the subjects are fully informed of this possibility and volunteer in spite of it.

Some current experimental research does not follow these principles. The subject is not always treated with the respect he deserves.

Black, Hillel, They Shall Not Pass. New York: William Morrow and Company, 1963, 342 pp.

Black, a skilled reporter, devotes his book to the thesis that many people in the United States are being penalized because of the test makers' overemphasis on the merchandising of tests and because of the ignorance of many school officials. Testing has become a way of excluding people rather than an aid to help children make the best possible choices for their future. There are some virtues to standardized tests: they can supplement personal judgments; they can have a beneficial influence on what is taught at particular schools; they can provide a universal yardstick that is particularly helpful in college selection.

But tests are doing incalculable harm to thousands of American children. Tests are imperfect measures of ability; yet they are merchandised with exaggeration and faulty claims. There are no ethical restrictions on these claims. School counselors and teachers using tests are woefully ignorant of the nature and purposes of the tests. Many cannot understand the language used to interpret them. Added to these inadequacies is the mystic faith in numbers that substitutes scores for the effort of trying to know each child individually.

In the college rat race, the competition for good test scores in order to enter the elite colleges becomes paramount. The admissions mania is the latest middle-class neurosis.

Personality questionnaires pose questions that could easily disturb the sensitive mind of a child. The test makers' descriptions of maladjustment are superficial and generalized. They try to force all children into the conformist mold.

There are things to be done about testing abuses. Parents should inform themselves about the limitations of tests. The public could urge formation of a Consumers' Test Bureau. All personality and career-choice questionnaires should be abolished from the schools. The public should urge the reduction of the number of admissions and scholarship examinations. There should be a crash program to train guidance counselors in the facts of measurement. Our attention should be devoted to <u>developing</u> talent by improving our educational opportunities rather than <u>searching</u> for talent with tests that too often fail to locate it anyway.

Bonner, John T., "A Biologist Looks at Unnatural Selection," <u>Princeton Alumni Weekly</u>, November 23, 1962, pp. 6-8, 16.

Pressures of population increase are manifesting themselves in increased pressure for admission to college. The decreasing proportion of applicants who are admitted suggests that the bases for college selection will have even greater implications for society in the future. A parallel with biological selection is drawn, in which the "characters" for which admissions officers are selecting students are examined and criticized. The author finds most of the items of admissions data -- objective test scores, high school re-

cords, and interviews -- quite deficient. He goes on to cite examples of individuals, such as Charles Darwin, who would be judged to have inferior credentials by current admissions standards. Although advocating bold and persistent research on the problem, the conclusion is reached that the uniqueness of each individual defies measurement, and that success or failure depends more on the "inner self."

Brim, Orville G., Jr., "American Attitudes toward Intelligence Tests," American Psychologist, Vol. 20, February 1965, pp. 125-130.

In 1963 the Russell Sage Foundation began a program of research on the social consequences of standardized ability tests. Results of two opinion surveys, one from a national sample of 1,500 adults and the other from a national sample of 60 secondary schools, provide some insights about anti-testing sentiment. Five issues are involved: inaccessibility of test data, invasion of privacy, rigidity in use of test scores, types of talent selected by tests, and fairness of tests to minority groups.

Most secondary school students believe they should be told their ability test scores, but neither they nor their parents are getting the information. Testers are afraid of possible misinterpretations, but steps must be taken to establish a collaborative relationship between tester and respondent in which both gain information of value to them.

Test data have become a key part of the new concept of a career record that accompanies a person throughout life. But who is to keep the record, and who is to have access to it? The criticism of tests as an invasion of privacy is directed more to tests of motive, beliefs, and attitudes than to tests of intelligence. The fact is that confidentia-

lity of personality test data cannot be protected. Test results are subject to subpoena by any group with proper legal authority and can easily become a matter of public record. Legally, under what conditions can the state invade an individual's right to privacy? Morally, is it sufficient to justify the asking of questions because of the eventual contribution to knowledge, on the assumption that the growth of knowledge about social sciences is a public good?

Rigidity in the use of tests makes no allowance for possible changes in the person or his future environment. It happens that the public believes that intelligence increases throughout life. Within this concept of intelligence, there is inevitable antagonism to the use of intelligence tests. It might help to give the public some education about the nature of intelligence tests -- that they do not measure wisdom as such. But, on the other hand, there is no doubt that the application of test results in many schools and other settings is much too rigid. We need provision for continuous appraisal of an individual's performance after he has been allocated to one or another environment. Though we can predict success, we must not treat possibilities as certainties.

Some oppose tests because they feel tests deny opportunity to persons with different and possibly highly valuable talents. The opportunity structure in American education and to some extent in American occupations is organized around intelligence tests. Creativity, ambition, honesty, altruism, and other important qualities are not measured by the test.

Interestingly enough, minority groups seem to be favorably inclined toward the use of ability tests because the tests constitute a universal standard of competence and potential. A comparison of Negro and white adult respondents in the Russell Sage study showed that at the lower social class

levels, Negroes had more favorable attitudes toward the use of tests in job selection and promotion than did white respondents.

What are the basic sources of these criticisms? First, there is opposition arising from some general personality characteristics; second, from systems of values. Third, antagonism develops as a consequence of an individual's experience with intelligence tests. Fourth, opposition arises from the restrictions on life opportunity that result from poor performance on tests.

The study uncovered one curious finding. People are apt to raise their intelligence estimates, no matter what kind of information they receive. This suggests a selective use of information designed to protect one's self-esteem, in which those who receive data that upgrade their ability estimates remember it and use it, and those who receive the contrary forget it or explain it away. The residue of displeasure may well remain and be directed into resentment against tests.

Brown, Spencer, "Gateway to the Colleges: An Examination of the College Entrance Board," Commentary, Vol. 27, June 1959, pp. 472-482.

To try to predict the nebulous concept of success in college by precise examinations is to measure a fogbank with a yardstick. The apparent precision of Board scores vanishes when we try to find what the scores are based on. It seems that the Board is dominated by statisticians who are no longer influenced by the teachers whose servants they ought to be.

There is no serious criticism of the reliability of the Board tests, but of their validity. Predicting success in

college is hazardous, and test makers themselves agree that college success demands more than a good test score. Yet the test makers insist that tests do measure something and that the correlation proves it. The Board's publicity fosters the notion that we need only look at the good correlation.

The English Composition Achievement Test is fundamentally illogical, no matter what its correlation with English grades. English teachers do not believe that the knack of editing copy and of compressing clauses into single phrases or single words is the quintessence of the art of writing. The difficulties of reading 100,000 essay examinations in English are not insurmountable if the Board were interested in overcoming them. Both college and high school teachers want more writing for their students.

Any examination system has a life of its own and becomes in some measure independent of the forces that created it. The College Board, in spite of its official disclaimer of responsibility for the high school curriculum, becomes more influential every day. It is the academic embodiment of mass civilization. Unfortunately, there is no assurance that colleges recognize the fallibility of the Board tests or that they realize the vast distinction between education and testing.

Campbell, Joel, "Testing of Culturally Different Groups," Research and Development Report 63-4, No. 14. Educational Testing Service, 1964, 22 pp.

The investigation was undertaken to examine data from several studies concerned with predicting the performance of Negroes and others from deprived backgrounds. The following major conclusions were drawn: cultural deprivation will affect test performance adversely; remedial efforts

can improve test performance, although the limits of this improvement have not been established; tests of verbal and arithmetic ability are effective predictors of academic grades in both white and Negro colleges.

Campbell, Roald F., Cunningham, Luvern L. and McPhee, Roderick F., The Organization and Control of American Schools. Columbus, Ohio: Charles E. Merrill Books, Inc., 1965, 553 pp.

This is a book on the many influences that shape our school system; one of these is the College Entrance Examination Board. A study of the impact on secondary schools of four national programs (those of the National Science Foundation, the National Merit Scholarship Program Corporation, the National Defense Education Act, and the College Entrance Examination Board) revealed that these programs were having a decided impact on secondary schools. The programs tended to reinforce each other and were chiefly concerned with the college preparatory function of the high school. There was some evidence that national programs tended to produce standardization in secondary school curriculums across the nation. While none of these programs was legally imposed on local school districts, it was hard for local districts, particularly suburban districts, to resist them. In a sense, acceptance of the programs tended to shift decision making from the local to the national level.

As one example, teachers do attempt to teach to tests. Concrete evidence on this point has been provided by Henry Brickell's New York State study. He reported that beyond any doubt, the Regents examinations inhibited change in the state of New York. Not only did the schools explicitly teach on the basis of previous examinations (copies of previous

tests constituted at least 10 percent of the curriculum in the high school course), but the schools tended to shy away from innovation because of the fear that the test record would suffer.

Concern for the impact of the external testing programs led three major national associations in education, the American Association of School Administrators, the Council of Chief State School Officers, and the National Association of Secondary-School Principals to establish a committee to observe the impact of tests on the secondary schools. In 1962, the committee expressed a fear of control of curriculum through testing. Their survey of school administrators showed that 70 percent of the respondents believed the tests were based on some concept of what <u>should</u> be taught. Nearly one-half said they used test results to aid in curriculum change and evaluation and to determine the extent to which teaching objectives were being attained. Using test results in this manner assumes that the objectives of the test makers are the same as those of the school system or of the teacher. It is a strange fact that those responsible for the construction and administration of the tests are among the most critical of the teachers who teach for the test.

In a dissertation on the influence of 10 national programs, including the College Board, on the curriculums of 11 selected independent secondary schools, Roy Larmee found that the greatest single influence on the curriculum policies of the schools was the set of course descriptions prepared for the Board's Advanced Placement Program. All of the schools in the study tried to prepare students to take these examinations, and satisfactory passage of one of the tests became a curriculum goal in each of the schools.

"It may very well be that these pressures from the colleges are all desirable and will result in improved educa-

tion across the country. That is not the issue here. The issue is that these changes are almost totally nonlocal in origin; rather, they originate primarily because the colleges are constantly seeking better-prepared students. The concept of guidance at the seventh-grade level for advanced placement in college may strike some as having overtones of 1984."

Carlson, Robert O., "The Issue of Privacy in Public Opinion Research," Public Opinion Quarterly, Vol. XXXI, Spring 1967, pp. 1-8.

Public alarm over invasion of privacy in social research may be expected to extend to public opinion polls, especially to the interview situation. It is necessary, for practical resons alone, to consider what benefits individuals receive from claims on their time and privacy that are made in connection with public opinion research. An extensive educational effort should be undertaken to explain the role of survey research in assisting planning within government and other agencies. Serious consideration must also be given to ways of making public opinion research contribute more concretely and visibly in our society.

Chauncey, Henry, and Dobbin, John E., Testing: Its Place in Education Today. New York: Harper and Row, 1963, 223 pp.

In general, this book would constitute a good defense of testing. It gives a lucid picture of tests and their place in the teaching and learning process. It covers the history of testing; tests of learning ability; achievement testing; tests as tools in teaching; tests in selection, admissions, and guidance. It asks questions about what characteristics

make a good test. It also gives some specific examples of various types of multiple-choice tests. It is not a crusading book, but it provides a very clear exposition for the layman as well as the educator.

Chauncey, Henry, and Hilton, Thomas L., "Are Aptitude Tests Valid for the High Able?" Science, Vol. 148, June 4, 1965, pp. 1297-1304.

Evidence from several studies refutes the allegation that aptitude tests are not valid for students of superior ability.

Associations between aptitude tests such as the Board's SAT, the Graduate Record Examinations, and the Miller Analogies Test, and criterion measures such as grade-point average, ratings of scientific accomplishment, number of Ph.D.s at a given aptitude level, American Men of Science and Who's Who listings, indicate that aptitude tests are valid predictors of various performance criteria for samples of individuals high in ability.

Test ceilings can be sufficiently high to discriminate among students of high ability.

Furthermore, there is no evidence that objective tests discriminate against superior students who are able to perceive imperfections in the keyed answers. If such a bias penalizing the high-ability student exists, it is small enough not to be detected in large samples, although the possibility that it may account for a few inconsistencies cannot be discarded completely.

Clark, Kenneth B., "Intelligence, the University and Society," The American Scholar, Vol. 36, Winter 1966-67, pp. 23-32.

Man has sought to understand the mysteries of his environment by asking questions about his origin and the meaning of his existence. The critical question of this period of human history is whether human intelligence as traditionally defined offers any reliable assurance of human survival. Though human intelligence and its richest consequences -- science, technology, art, literature, philosophy, and religion -- are essential to survival, they do not in themselves reduce the capricious dangers to human existence. Modern man's ignorance lies in an inadequate functional sense of social morality.

Our universities must produce human beings with morally sensitive intelligence. Yet our universities try to escape this role. In fact, they have had a long history of default on important moral issues. They have tried to make a virtue of isolation from daily social problems. They have claimed that academic detachment and scientific objectivity are their tools, and, thus escaping from value commitments, they contribute to moral erosion.

The universities have facilitated moral emptiness by supporting the process in which education from the primary grades on has become ruthlessly competitive and anxiety-producing. Schoolchildren are taught that intelligenc is the way to attain superior status and economic advantage over others. "Under the guise of efficiency, the demands of mass education and the pressure of limited facilities in colleges, the schools have facilitated the reduction of the educational process to the level of content retention required for the necessary score on the College Boards and the Graduate Record Examinations at the price of reflective and cri-

tical thought."

American higher education need not continue subordinating itself to the goals of efficiency, expediency, power, status, and success. It can instead produce totally educated people who value independent thought, individuality and creativity, concern and social commitment.

Clark, Kenneth B., and Plotkin, Lawrence, The Negro Student at Integrated Colleges. New York: National Scholarship Service and Fund for Negro Students, 1963, 59 pp.

This booklet describes the findings of a five-year follow-up study of the National Scholarship Service and Fund for Negro Students. The subjects were the 1,519 students who, as high school seniors, sought some type of aid, counseling, or financial assistance from the Fund in order to enter interracial colleges in the years 1952 to 1956. Complete information was available for 509 of these students.

An outstanding finding was the relatively low drop-out rate of this group of students, about one-half the national average for whites and Negroes at segregated colleges. Yet the predictive value of precollege test scores was not high, in terms of college grades. The study indicates that motivational factors are probably more important than test scores in the demonstrated superiority of Negro students in completing college. The authors recommend, in fact, that college admissions officers weigh test scores less heavily for these students since they do not predict college success as they do for white students.

The low drop-out rate cannot be explained in terms of superior academic performance by the Negro students. There was a marked relationship between high school average and academic success in college. Negro college students were below

the total college population on the SAT. "To rely on the alleged predictiveness of test scores in evaluating these students would ignore major findings of the study and exclude many capable students from college."

Dressel, Paul L., "Testing in Retrospect and Prospect," Nineteenth Yearbook of The National Council on Measurement in Education. Ames, Iowa: The Council, 1962, pp. 64-66.

"The answer to the lament, 'tests, tests, tests,' is not fewer but more, better, and better coordinated evaluation." It seems to be quite true that since 1940 there has been marked increase in the use of tests in placement, counseling, and in the appraisal of education at all levels. In a sense, this increase in testing is almost inevitable. It is essential that in a democracy each individual receive the education that will most fully develop his potential and that he be so placed that he simultaneously contributes to the society and obtains a high degree of personal satisfaction. Assessments of individual potential are complex and fraught with error. But certainly judgments made by a single individual are more likely to be in error than the composite judgments of several individuals. In a democratic society, every form of appraisal will have its critics, and it is well that this is so, for continuing modification of appraisal practices is always necessary in a dynamic society, and improvement is always possible. Criticism is a spur to change, to improvement, and to the development of procedures for appeal of incorrect appraisal.

As we move toward greater uniformity in standards of all levels of education, we shall need ways of assessing the level and progress of an individual as well as the quality

of an educational program. At the present time, our means of institutional accreditation are relatively crude and provide no assurance of educational quality. We owe it to our young people to develop some means of informing them of the return on their investment.

The means of appraisal must always include the possibility of appeal or repetition in order to provide an adequate safeguard for borderline decisions. The extent of error, the number of errors that can be tolerated, and the seriousness of a misjudgment must all be considered in deciding on the appraisal system.

Dressel is convinced that the development of objective, widely used standards of appraisal is absolutely vital to development of our democratic society. Where the criteria and the means of appraisal are covert, there can be no assurance of justice and no assurance of improvement.

Ebel, Robert L., "The Social Consequences of Educational Testing," Proceedings of the 1963 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1964, pp. 130-143.

In recent times, testers have been charged with showing lack of proper concern for the social consequences of our educational testing. There are four themes to the criticisms: (1) Educational testing results in permanent status determination. It predetermines the adult social status and does irreparable harm to the self-esteem. (2) Educational testing can lead to a narrow conception of ability and reduce the diversity of talent that is available to society. (3) Testers can be in a position to control education and determine the destinies of individuals while incidentally making themselves rich in the process. (4) Educational

testing may encourage inflexible, mechanistic processes of evaluation and determination.

Instead of trying to dispel these apprehensions, the author decides to accept them as having some basis in fact and sets himself the task of discovering those things that might be done to limit the causes for concern. With regard to permanent status determination, most testers are well aware of the fact that there is no direct, unequivocal means for measuring permanent general capacity for learning. Intelligence tests are direct measures of achievement in learning, including learning how to learn, and inferences from these scores to some native capacity for learning are fraught with hazards. But the layman does not know this. Test specialists discredit the popular conception of the IQ and suggest that talent is something that can be educationally developed. It is better to emphasize the opportunity for choice and the importance of effort than to stress genetic determinism of status and success. This means that tests should be judged not in terms of how accurately they enable us to predict later achievement but how much help they provide in increasing achievement by motivating and directing the efforts of students and teachers. The author recognizes that many psychologists would not agree with this definition that the immediate purpose of measurement is always description, not prediction or control.

The danger that a single widely used test may foster an undesirably narrow conception of ability is not completely imaginary. The problem of encouraging various kinds of ability is much broader than the problem of testing. But perhaps those who manage testing programs should permit variation in the test administered from one person to another. The use of optional tests of achievement is one way of accomplishing this. It is convenient to use a common yardstick, but this means that some students with

special talent are neglected.

If the test maker persists in secrecy about tests and test scores, the general public will fear that the tester has too much control over students' destinies. The essential information revealed by test scores could be communicated to laymen. It is true that test scores can be misused by the layman, but this does not justify the withholding of knowledge. Nor can we overlook the practical reason for secrecy regarding test scores -- that is, it spares those who use the scores from having to explain and justify decisions tney make. If decisions cannot be justified, perhaps tests ought not to be used as components. Testers do not control education, and by the avoidance of mystery and secrecy they can help to create better understanding and support.

Tests should be used as little as possible to impose decisions and courses of action on others. They should provide a sounder basis of choice in individual decision making. There are no universally accepted goals of human behavior.

What are the social consequences of not testing? Only to the degree to which educational institutions can define what they mean by competence and determine the extent to which it has been achieved can they discharge their obligations to society. If tests were abandoned, encouragement and reward of individual efforts to learn would be more difficult. Educational opportunities would be extended less on the bases of aptitude and merit, and more on the bases of ancestry and influence. Decisions on curriculum and method would be made less on the basis of solid evidence, and more on prejudice and caprice. In Ebel's judgment, these social consequences are potentially far more harmful than any possible adverse consequences of testing.

Findley, Warren G., "Testing's Second Chance," Twentieth Yearbook of the National Council on Measurement in Education. East Lansing, Michigan: The Council, 1963, pp. 70-74.

There are three serious problems in testing to which careful attention should be paid. (1) Testing programs should not inhibit new developments in curriculums. Therefore, they should be revised frequently. A related danger is confining tests to those areas which are not affected by curricular change. (2) Test results should not be used exclusively as tickets of admission to high-prestige colleges. They should also be utilized for vocational guidance and instruction. Students who fail to qualify for admission to college should be trained and given a second chance. (3) Tests should not be administered haphazardly with no specific purpose in mind. Few persons are well trained in the construction and use of tests. There is the danger that untrained counselors will misuse tests.

Fishman, Joshua A., and Clifford, Paul I., "What Can Mass-Testing Programs Do for-and-to the Pursuit of Excellence in American Education," Harvard Educational Review, Vol. 34, Winter 1964, pp. 63-79.

Mass testing is interwined with the functioning of the American educational system, and it is under strong and continuous attack by those who see it as undermining the quality of education programs. However, much of the criticism is misdirected because of an inadequate perception of the diverse roles and functions of mass testing. Serious limitations exist in testing -- not the least of which is the relative inability of such programs to break out of traditional patterns. Nevertheless, the mechanization of educa-

tion that many fear cannot be attributed to testing per se; rather, what is needed is a fundamental reexamination of our educational goals and methods.

Fiske, Donald W., "The Subject Reacts to Tests," American Psychologist, Vol. 22, April 1967, pp. 287-296.

Perceptions of tests and reactions to tests were studied through interviews with a representative national sample of 589 adults. The study indicated that subjects are not completely in the dark about the purposes of tests and their levels of utility. Subjects do react in different ways to personality tests, because the typical personality test is not precisely defined. This diversity of reaction undoubtedly contributes to the variation in obtained scores. Tests of ability, on the other hand, present the subject with a clear task which he understands and is willing to perform. In a testing situation the subject reacts to the knowledge that he is being evaluated. Both personality tests and ability tests present potentially upsetting stimuli. The effects of reactions on test responses need study.

Subjects manifest a wide variety of reactions to tests and to being tested. Some think tests are good, others that tests are worthless. Some are bored, some are highly involved. Differential reactions are to a small extent associated with education and exposure to views about tests.

Fricke, Benno G., "Review of the College Entrance Examination Board's Admissions Tests," The Sixth Mental Measurement Yearbook, edited by Oscar Buros. Highland Park, New Jersey: Gryphon Press, 1965, pp. 975-996.

This review gives an overall evaluation of the admissions test package.

"Unfortunately and surprisingly, while most of these tests have been in existence for over twenty years, there is relatively little research evidence on which to base a judgment." There is no test manual or handbook containing the results of research, and the Board particularly needs uncontaminated research results from institutions that do not rely on test data in arriving at admissions decisions.

Though it should be possible to construct subtests that measure verbal and mathematical aptitudes quite separately, there is a disturbingly high correlation between scores on the verbal and the mathematical sections of the Scholastic Aptitude Test. Though the two tests are not measuring the same thing, they overlap to an undue extent. More important than their intercorrelation is the evidence that the two scores have similar correlations with both appropriate and inappropriate subjects in the college curriculum. That is, grades in science and mathematics courses are not predicted significantly better by the mathematical sections than by the verbal sections.

There is very little information on the specific or distinctive validity of each of the Achievement Tests. The writer carried out his own studies on freshmen at the University of Michigan and found that scores did not correlate appreciably higher with the courses they were supposed to predict than with other courses. Each Achievement Test seems to be mainly a measure of general ability. It also appears that the SAT and Achievement Tests are measuring similar if not identical abilities.

Probably the major reason for the high correlation between the various tests and for their lack of validity is to be found in the item analysis procedures used to construct new forms. This procedure utilizes an internal rather than an external criterion to determine which items are to be selected for new forms. This encourages considerable in-

breeding of test items. Scores from these tests are homogeneous and highly reliable but not highly valid for external criteria such as college grades. It would appear that reliability has been stressed at the expense of validity. An alternative procedure would be to use grades in college as the basis for the item analysis.

In general, the face validity of test questions is excellent. But for students who have had a relevant course in high school, each Achievement Test functions mainly as a general academic ability test. This is a serious matter, and it has led to excessive weight being given to test ability and, probably, test-wiseness. It would be far better to have one valid ability test score.

In Fricke's opinion, more harm than good results from the use of Achievement Tests that are not good measures of what they purport to measure.

Gardner, John W., Excellence: Can We Be Equal and Excellent Too? New York: Harper Colophon Books, Harper and Row, 1961, 171 pp.

The chief instrument used in the search for talent is the standardized test. Not surprisingly, tests have been the subject of considerable hostility. For some, the aversion to tests is defensive: they fear precise appraisal of their own or their children's capacities. For others, the aversion is simply a normal reaction to what they consider as invasion of privacy. Some fear the tests will come up with an unfair appraisal. Reassurance about high statistical reliability and validity does not help much. Apprehension is fostered by the fact that it is very hard for the nonprofessional to understand mental measurement. No one wishes to be judged by a process he cannot comprehend.

To some degree, anxiety about tests is a fear of the potentiality for social manipulation and control inherent in any large-scale processing of individuals. There is not only fear of the tests themselves but of the unknown bureaucracy that handles the test and acts on the results.

Yet probably if these sources of concern were dissolved, the hostility toward tests would remain. Tests are designed to do an unpopular job. It happens that tests are excellent when limited to the use for which they were designed. The development of standardized tests is one of the great success stories in the objective study of human behavior. Although it is now said that tests give an unfair advantage to the privileged individual, before tests many people seriously believed that the less-educated segments of society were not capable of being educated.

Anyone who attacks the usefulness of tests must suggest workable alternatives. At the present time they have proved fairer and more reliable than any other method when they are used cautiously. The best achievement and aptitude tests are remarkably effective in sorting out students according to their actual and potential classroom performance.

Of all mistakes in applying tests, perhaps the worst is in extending them beyond the strictly academic or intellectual performance for which they were designed. Everyone knows that there are other important ingredients in success -- aptitudes, values, motives. The youth who has zeal, judgment, and staying power may not be selected in school as a person with high potential, but he may earn marked success in later life.

Some rules can be suggested for minimizing the hazards and maximizing the benefits of tests. First of all, they should not be the sole way of identifying talent. A second rule is that the diagnosis of aptitude and achievement must be a continuing process. It is not enough to say that

a child has been tested; he must be tested consistently over the years. We do not accept a test score that is several years old any more than we accept a health report of a similar vintage. In these repeated testings, we expect aptitudes to remain pretty stable, but the fact is that at any given age level, a test score may not be a precise reflection of aptitude. And also, the student himself may change from year to year -- if not in aptitude, then in achievement, motivation, and many other crucial dimensions.

There are many socially valuable kinds of talent not measured by aptitude and achievement tests. Although this sounds obvious, the easiest and laziest thing to do is to sort youngsters out by aptitude scores and forget the rest. The sorting of individuals in a society is an exceedingly serious and explosive business. Because the consequences for the individual are so serious, the final weighing of evidence must be made by a qualified and responsible human being rather than a machine. It is tempting to place complete faith in the rapid and efficient handling of large numbers of individuals through scores. But considerations of efficiency must not narrow our conception of talent.

Goslin, David A., "The Social Impact of Testing," Personnel and Guidance Journal, Vol. 45, March 1967, pp. 676-682.

Although problems remain in developing more valid tests and in improving the use of tests for counseling and selection, the time has come when attention should be directed to a "second generation of testing problems." These concern the social consequences of continued widespread use of tests, in terms of their impact on the individuals involved and the groups that use them. Some of these questions are: What

is the objective influence of tests on the opportunities open to individuals? What part do test scores play in influencing the kinds of advice given to young people in a variety of situations? Does "objective" information about ability have special effects on the opinions a person holds about himself? What would be the ultimate effect on society of a wholehearted commitment to tests as the means of evaluating abilities of individuals?

Survey data collected as a part of the Russell Sage Foundation study reveals that counselors and teachers are unaware of the extent to which they make use of test scores, and that there is ambivalence about the dissemination of scores to pupils and parents. Further, Rosenthal and Jacobson have found that teachers' expectations of intellectual growth are associated with the actual amount of improvement in test performance which is subsequently observed, even when the subjects so identified are a randomly selected group.

The implication of these findings is that much more attention must be given to the problems of access to scores, and to reduction of the self-fulfilling prophecy that such information may produce. The dimensions of the problem of score dissemination and use are manifold, requiring attention not only to experimental variables, but also to policy questions that touch on legal rights and ethical considerations.

Gross, Martin L., The Brain Watchers. New York: New American Library, 1963, 256 pp.

Gross, a professional writer, devotes the major part of his book to the damage done by the brain watcher or personality tester. The book is liberally sprinkled with quotations

from testing officials as well as from professional journals and books.

Personality testing is a nonscience. The moral implications of personality testing lie not only in its inaccuracy but its approach to group statistical guilt. A man is accused and convicted for his variance from a norm. The tester is in the service of management, and his compulsion to protect what he thinks are the best interests of the corporation rather than the individual creates many of the dangers. Though the immorality of false prediction is obvious, it may also be that the mere attempt to predict the behavior of individuals is a violation of personal destiny. The prediction influences the subsequent events. While the battle against discrimination in employment has been proceeding reasonably well, the subtle discriminations of personality testing have not been adequately recognized, much less removed. The individual is deemphasized in the interest of establishing a safe hiring policy that will not agitate the management.

But corporations, government agencies, and factories are not the only institutions at fault. There is brain watching in our schools. The tester has been working furiously to find a statistical correlation between personality and grades so that he can tap a vast new commercial market. The College Entrance Examination Board has given research grants to psychologists to find personality types slated for college academic success. If this search is successful, tests will be used to predict which high school seniors have the personality makings and which should be rejected because they have failed their "personality boards." Unfortunately, the brain watchers are already at work on college campuses screening out those with undesirable personalities.

The brain watcher in the school is most often the naive guidance counselor, an individual without professional train-

ing in the interpretation of personality tests. The innocent child, trusting his teacher, will answer any personality question that he is asked. What checks have been made to make sure that the student is not being upset? What validity data exist for these tests?

Though the College Board's tests are not brain-watching tools in the strictest sense, the brain watchers are headed toward some sort of personality testing for college entrance. The ridiculous idea of mating student and college personality has been taken quite seriously by the College Board, which has given financial support to research on the idea through the College and University Environment Scales.

It is quite true that there are many sober studies of personality tests. But one conclusion is clear: they show complete and chaotic disagreement. The inability to find the measurable links between intellect and personality upsets the tester. No one personality has any monopoly on good college grades or intellectual accomplishment. The testers should know this.

Hathaway, Starke R., "MMPI: Professional Use by Professional People," American Psychologist, Vol. 19, March 1964, pp. 204-210.

Tests do not invade an individual's privacy. Items on the Minnesota Multiphasic Personality Inventory related to religious activities were not constructed to inquire about the particulars of one's religious beliefs but to identify psychiatric disturbances, the symptomology of which often involve certain patterns of religious expression and thought.

The items dealing with religiosity are not evaluated individually but in combination with other items that do

not seem to relate to religion. The individual's total score on a dimension is taken into account in evaluating his health or maladjustment, not his specific response to a single item, which is seldom scrutinized by anybody.

The MMPI serves as an effective screening tool. Its judicious use can help protect the person being hired as well as the person hiring him.

To make sure that those who want to can preserve their privacy, examinees should be informed that they may omit any item they do not wish to answer for any reason.

Hawes, Gene R., "Knowing the Score," Columbia University Forum, Vol. VIII, Spring 1965, pp. 47-49.

Psychological testing should not be considered an invasion of privacy falling into the same category as wire tapping and bugging. An individual is not spied on without his knowledge. The information is collected with the cooperation of the person being tested, although his motive for cooperating may be to avoid a less pleasant alternative. Moreover, the results of psychological tests are customarily treated as confidential

However, if the individual is not told what a certain test is for, he may unwittingly be revealing information about himself that he may not wish to divulge. Therefore, apprising the person of traits measured by the tests would allow him to exercise the same degree of control he normally has over application blanks and interviews. In addition, if test results are released to the examinees and their use explained, most objections to testing as an invasion of privacy should disappear.

Current objective ability and achievement tests are quite useful and result in consistently higher correlations than

any of the other methods of selection. Personality tests, however, lead to statements of extremely low probability and produce an inordinately large number of false positives and false negatives. They can also be easily faked. The use of personality questionnaires, therefore, can be challenged on these grounds.

Criticisms of testing should be constructive to inspire progress. Exaggerated caricatures of testing are not particularly useful.

Hoffmann, Banesh, "'Best Answers' or Better Minds?" The American Scholar, Vol. 28, Spring 1959, pp. 195-202.

Examples of several multiple-choice items are cited, and the argument is offered that the gifted students would, because of superior knowledge and greater originality, choose what the test constructor regarded as an incorrect alternative. Items contained in the College Board SAT are frequently defective in this sense, containing numbers of items that are traps for the superior student and rewarding to the superficial one. Because multiple-choice tests generally are replete with ambiguities, and because they favor the superficially brilliant and punish the creatively profound, they exert a "baleful influence on teachers and teaching." Their widespread use is cause for grave concern, and there is a serious need for a full-scale inquiry into the whole field of testing. Such an inquiry might draw on representatives from distinguished scholarly organizations and might lead to the policing of testing or to the setting up of alternate systems of testing.

Hoffmann, Banesh, "The College Boards Fail the Test," The New York Times Magazine, October 24, 1965.

The College Board tests have been constructed with elaborate professional care, but their evolution has been molded by statistics. One of the basic defects of multiple-choice tests is that they ignore quality and are concerned only with the choice of answers, not with the reason for these choices. Objective tests insist on conformity, refuse to let the student express himself in words, exclude evaluative judgment. Test questions themselves contain many ambiguities.

A closer look at the testing process reveals that it is more scientism than science. It is based on statistics and therefore on the assumption that we can reduce important appraisals to numerical terms. Excellence, in the deepest sense, is not likely to be discovered by statistical techniques.

Had the Board believed its own statistical arguments, it would have substituted the SAT-verbal test for any of its English tests. But it finally had the courage to undertake research leading toward the inclusion of an essay in the testing program. There is a ray of hope here. There is a chance to break through the multiple-choice barrier.

Hoffmann, Banesh, "Psychometric Scientism," Phi Delta Kappan, April 1967, pp. 381-386.

There does not presently exist any generally satisfactory method for evaluating human abilities. Current attempts based on techniques of mass production and on the psychometricians' misuse of statistics are not only dangerous but, in a profound sense, unscientific. Arguments and examples are offered to support the position that pretest statistics are misleading and inherently so, for they suffer from the defects of multiple-choice tests themselves. For example, only exceptional students are apt to

see the deeper defects of test items, and, since these students are in a minority, item statistics are not sensitive to their presence.

The fallacy of the statistical criterion is that it necessarily must refer only to those criteria that are quantifiable. Further, it ignores the side effects of tests and thus serves to corrupt the educational process.

The incidence of defective items is far higher than generally supposed. But perhaps equally important is the response of psychometricians to criticism, where defensiveness and lack of objectivity have characterized their attitudes.

Finally, recent experiments on computer grading of essays display most clearly the glaring inadequacies of a mechanized approach to evaluation of writing ability and the inability of statistical evidence alone to reveal these shortcomings.

Hoffmann, Banesh, The Tyranny of Testing. New York: Crowell-Collier Press, 1962, 223 pp.

Objective tests are grossly unfair and inadequate. Defective questions are abundant even in well-constructed aptitude tests. Items are awkwardly stated; distractors are incomplete in detail; often, more than one alternative is equally correct. Unfortunately the "statistical magic" of the test constructor is an effective smoke screen.

Objective tests measure one's ability to answer trifling questions. They favor the superficial and cynically test-wise student. They penalize intellectually honest, individualistic, probing, creative, and superior students. Such tests stifle creativity, prohibit the student from explaining his choice, penalize the student who knows too much, favor conformity, mistrust individual judgment,

and foster intellectual dishonesty and opportunism, all of which warp our sense of values.

Objective tests are not worthy of first-rate minds. They fail to measure important ingredients of greatness.

Measuring English composition by objective items is outrageous. In fact, deterioration of English composition in secondary schools can be traced to College Board tests, as meaningful work in writing is being abandoned in favor of vocabulary drills, which can earn the student a high score on aptitude tests.

The shortcomings of essay tests enumerated by psychologists sound reasonable and logical, yet they feel wrong. The nonquantifiable aspects of testing should not be ignored, but encouraged.

Hoffmann recommends a committee of inquiry to examine the quality of multiple-choice tests and their makers.

Krasner, Leonard, "Behavior Control and Social Responsibility," American Psychologist, Vol. 17, April 1962, pp. 199-204.

While the issue of behavioral control first arose with regard to psychotherapy, it is now far broader and covers other areas, such as operant conditioning, teaching machines, hypnosis, sensory deprivation, subliminal stimulation, and similar research. There is considerable public interest, concern, and misunderstanding about the range and power of psychological findings.

A "psychology of behavior control" would differ from the science of psychology in subtle but important ways. The science of psychology seeks to determine the lawful relationships in behavior. In a "psychology of behavior control," these lawful relationships are used, deliberately,

to influence, control, or change behavior. This implies a controller and, with it, an ethical and value system of the controller. Because science is moving at a very rapid pace, now is the time to concern ourselves with the matter of control.

Two major steps are suggested. The first is to develop techniques of approaching experimentally the basic problem of social and ethical issues involved in behavior control. A second major step is communication between the general public and research investigators. Researchers must keep in contact with each other, and their work should be open to the public. It is the psychologist-researcher who should undertake the task of contact with the public rather than leaving it to sensationalists and popularizers.

Psychologists have no choice but to continue their research into human behavior. The danger is not in the findings but in their potential misuse. Safeguards can be incorporated into this type of research by deliberate recognition of the facts that the psychologist can influence other people's behavior and that this implies a value decision as to what is good behavior, what is mental health, and what is desirable adjustment. The fact that the behavior controllers are professional individuals is no guarantee that behavior control will not be misused. We have only to consider the role of German physicians in wartime medical atrocities as evidence of misuse by a supposedly professional group. Awareness is a major ingredient in defense of manipulation.

Malcolm, Donald, "A Summary of Criticisms of Aptitude and Achievement Testing which Have Appeared in Recent Newspaper and Magazine Articles." Memorandum from ETS, April 3, 1961.

This paper reviews 24 articles on testing and summarizes the objections to multiple-choice testing programs. The articles appeared between 1957 and 1961, either in newspapers or national magazines.

Comments are summarized as follows: (1) Test making is big business. (2) Tests endanger freedom of choice of individuals. (3) Tests have adverse effects on school programs: local control is lost. (4) Tests have an adverse effect on classroom teaching because teachers neglect the valid objectives of instruction. (5) Tests encourage students to take courses in which they will make good scores, create anxiety in students, and do not reflect their true abilities. (6) Tests have inherent defects -- they can measure only limited aspects of behavior, they oversimplify complex issues, they measure test-taking ability rather than real knowledge. (7) Teachers know more about aptitude and achievement of their students than one can learn from test scores. (8) Testers are not competent. (9) Testing is dominated by statisticians.

Mayer, Martin, The Schools. New York: Doubleday and Company, 1963, 499 pp.

Mayer is a reporter who spent several years in an extensive investigation of schools. He visited about a thousand classrooms in more than 150 primary and secondary schools in the United States, Britain, France, and Scandinavia. The general purpose was to wander amidst the current controversies over education.

In American schools, all children take at least one and most children take four intelligence tests -- usually in grades one and two and four and five, always in grade seven or eight, and often in grade ten. Generally, the influence

of the IQ score tends to be subtle rather than gross and to show up in expectations and in guidance. If a child scores high and does well, or scores low and does poorly, nobody worries about him. But if his work does not follow the path of the test score, the guidance people become concerned. Nowhere else in the world, except perhaps in Britain, does the IQ score influence people's expectations of the child as much as in the United States. Intelligence testing has created the fallacy that success or failure with a certain set of materials is governed only by the child's native aptitude. Tasks are set for various ages, and if students fail them, no one bothers to look at the teaching within the school, but only at the students.

Everyone accepts the fact that tests are class-biased. All the disagreement is on how the facts should be interpreted. That is, what should the schools teach and to whom? If the difference in test scores represents a true difference in the innate capacity of children from varying social backgrounds, then they cannot be pushed through a lengthy education. The argument has been continuous since the beginning of intelligence testing and was at its angriest during the nature versus nurture controversy of the 1930s. The central social problem in education is not that intelligence tests are biased, though they are, but that the schools themselves are biased.

The school administrator who needs to know how the children are doing has two choices: he can accept his teachers' opinions, or he can go looking for some outside measurement. When he goes out to raise tax money he likes to have the facts and figures about how well the children in his schools are doing in comparison with other schools. There are plenty of standardized tests at hand. It should be noted that it is the administrators, not the teachers or children, who are enthusiastic about these tests. "Indeed,

the Educational Testing Service, a non-profit group which runs the most 'scientific' test selling operation in the country, has found it necessary in its literature to proclaim that 'if a teacher is unsympathetic to a testing program, she is abdicating her rightful position.'"

The machine scoring aspects of objective testing are particularly repellent to some administrators, but they are certainly more just than the grades given to essays by individual teachers.

The construction of the standardized tests is a painstaking business. However, the main supporters of tests are those who make their living through testing. The usual objection is that the questions are bad. However, nobody has ever constructed a school system in which geniuses do very well. One should not condemn the tests simply because they reflect the limits of the schools. What is objectionable is the claim that they eliminate error. They simply move human error from the marking process to the test-writing process. It is not so much that they penalize the really bright child as that they paralyze the average child.

The most damaging criticisms of the tests deal with the assumptions that underlie them rather than with the details of item construction. For example, the desire to create reliable tests means a narrowing in the range of tasks that the child is asked to perform. By and large, scores on different achievement tests correlate too highly with one another. The tests have moved toward the measurement of those factors that affect high school grades, which in turn are related to college grades. This kind of reasoning assumes the accuracy of the grades given by teachers -- an assumption that would negate the necessity for standardized tests.

What is most distressing is not the inadequacy of tests as educational tools, but the literature the testers use to

promote them. ETS should set a standard of behavior for the field, but it is publicly committed to the proposition that testing is indispensable. It seeks to prove that standardized tests are the heart of guidance; it demonstrates the "need" for tests in the new curriculums; it has even begun speaking about creativity. The officers of a tax-exempt, charitably supported organization have a powerful moral obligation to keep their public statements well away from hard-sell salesmanship.

It is quite true that objections to multiple-choice tests apply equally well to essay or oral tests. That is, they, too, insist that every child must have learned the same thing; they contain class bias; they contain marking error. But it is the surrounding veil of science and publicity that make the objective tests so dangerous. In the real world, judgments are made by observations of people at work, not by the results of paper tests. If teachers were absolutely perceptive, there would be no need for tests.

Messick, Samuel, "Personality Measurement and the Ethics of Assessment," American Psychologist, Vol. 20, February 1965, pp. 136-142.

Public dissatisfaction with testing may be expected to lead to demands that personality assessment be sharply limited or controlled possibly through legislative action. Arguments for self-regulation by the relevant professions have thus far failed to deal with the conflicts in norms and values existing in the problems of assessment and regulation. The psychologist "believes in the dignity and worth of the individual," but "he is committed to man's understanding of himself and others." Policy decisions must be made in the face of a serious dilemma concerning these conflicting commitments.

Moughamian, Henry, "General Overview of Trends in Testing," Review of Educational Research, Vol. XXXV, February 1965, pp. 5-16.

In this survey of research on educational and psychological testing from 1962 to 1965, the author notes that tests and testing have been focal points of criticism. A common element in most of the criticisms had been the contention that far too many persons who give tests lack the ability to interpret test data.

Both professional and lay critics cited the inadequate preparation of teachers in tests and measurements. Recognizing the importance of this problem, a committee of the National Council on Measurement in Education prepared a test of teacher competence in educational measurement. If continued progress is to be made in testing, it is essential that marked improvement be realized in the interpretation of test scores.

Panel on Privacy and Behavioral Research, "Preliminary Summary," Science, Vol. 155, February 3, 1967, pp. 535-538.

This is a preliminary summary of the report of the Panel on Privacy and Behavioral Research appointed by the President's Office of Science and Technology. The chairman of the panel was Kenneth E. Clark, dean of the College of Arts and Sciences, University of Rochester. The panel was appointed to examine the issue of the invasion of privacy in behavioral research and to propose guidelines for those connected with this research.

The privacy problem in scientific research is small compared with that in employment interviewing, social welfare screening, and law enforcement investigation. Nevertheless,

there are instances in which behavioral scientists have not followed appropriate procedures to protect the rights of their subjects. Because of this, there has been pressure from some quarters, both inside and outside the government, to place arbitrary limits on research methods. This creates a conflict between two dominant values in American society. One is that the individual has an inalienable right to dignity, self-respect, and the freedom to determine his own thoughts and actions within the broad limits set by the requirements of society. The other is that the scientist is not to be hampered by restrictions. Science has the right to explore any part of the universe, including man. How can these values be reconciled?

In the end, it must be accepted that behavioral research will sometimes conflict with the principle of privacy. There must be constant weighing of the costs and the gains.

Behavioral science seeks to assess many aspects of men's minds and feelings. Without informed consent on the part of the subject, these measurements represent invasion of privacy. Yet the traditional concept of informed consent needs modification for certain types of behavioral research. There are situations in which the nature of the inquiry cannot be explained adequately or in which an explanation would invalidate the experiment itself. In these cases, the relationship between the subject and the scientist and between the subject and the institution sponsoring the scientist must be based on trust. The scientist and sponsor must protect the privacy and dignity of the subject. They must agree to treat the subject fairly and to cause him no inconvenience or discomfort unless this has been accepted in advance by the subject. Where even this degree of consent cannot be obtained (naturalist observations of group behavior, for example), the scientist has the obligation to ensure full confidentiality of the records.

Increased federal support of behavioral science has increased potential dangers as well as gains. Government must maintain the highest standards for the research it supports. Although the primary ethical responsibility rests with the individual investigator, governmental sponsors must be sure that both the investigator and his institution take the necessary steps to discharge their responsibility to the human subjects involved. Legislation on this is neither necessary nor desirable. The methods for institutional review can be determined by the institutions themselves, and research instruments should not be subject to detailed review by government funding agencies.

A set of recommendations on these points is presented.

Rychlak, Joseph F., "Control and Prediction and the Clinician," American Psychologist, Vol. 19, March 1964, pp. 186-190.

Control and prediction are often cited as the heart of any science, but the significations of these terms are not clear cut. The writer contends that there are three major uses of the phrase. The most legitimate would be as a theory of knowledge: the scientist designs experiments by controlling the empirically defined variables that seem to have most relevance to the hypothesis, and he predicts outcomes so that verified hypotheses can be taken seriously. Control can be a logical as well as literal activity of the experiment.

Second, prediction and control can be used as a language of description. Prediction would be an informal method of behavioral description in routine assessment statements. This is not predicting to validate an experimental hypothesis but is a more informal usage. It rests on the theo-

retical assumption of a universe of stimuli and responses all somehow bound together and therefore controlling and controlled.

Third, prediction and control represent a method of social influence. This is a usage that has ethical implications for many clinical psychologists. A clinical psychologist is not committed to accepting this ethical principle simply because research demonstrates that he influences his client to change behavior. But once he does, he must realize that he is saying that it is right that he, as a psychologist, with certain knowledge and training, makes desisions for others and consciously, deliberately, influences their behavior in ways that the research designates as good or correct.

Sorokin, Pitirim, Fads and Foibles in Modern Sociology and Related Sciences. Chicago: Henry Regnery Company, 1956.

Any science contains truth, half-truth, sham-truth, and plain error. The purpose of these essays is to display the nonscientific elements in modern sociology and psychology.

One chapter is devoted solely to "testomania," the process in which every individual is tested from the cradle to the grave, before and after the important events in life. The enormous influence of tests on the life-career is due to their supposedly precise and scientific nature.

The process of testing goes on incessantly in all differentiated, stratified, and long-living societies in order to be sure that members are tested and sorted into various social postions, strata, ranks, occupations, and activities. These "tests" take many forms. Some are the real and continuous institutional evaluations by the family, the school, the church, and social and occupational groups. Some are

life-tests of a man's ability to handle crises -- the commander of the battle, the champion in the ring, the contender for the throne. These tests are real but are short and sporadic. Some are longer tests of ability as revealed by a period of probation. Some are ad hoc, artificial, and magic tests, as old as history itself -- signs, rituals, conformations, and so forth.

Modern psychosocial tests are doomed because they try to measure the unmeasurable -- the fickle, unstable, and complex nature of man. Even real tests in life situations are riddled with errors; psychosocial tests are even more likely to blunder. Rarely do they involve actual performance; most often, they are short pencil-and-paper and vocal tests given sporadically under conditions decided by the testers and not the ones being tested. Consequently, the results have a chance character. Not everyone can answer instantaneously all sorts of questions. One must take into account temporary moods, styles, indispositions; but tests do not allow for these. They especially penalize those who mobilize their resources slowly. In addition, the very test questions themselves have inadequacies -- ambiguities, lack of single correct answers, stress on informational capital of the individual.

Unlike measurements in the physical sciences, psychosocial test scores have no meaning per se. They acquire meaning only when interpreted by the tester. These superimposed interpretations are rarely based on a proved causal link between test result and specific interpretation. Mainly they are derived from dogmatic belief in results as repressed wishes, native intelligence, true syndromes, and so forth. Thus the interpreter adds nonscientific elements to test scores. The result is invalidity.

The wondrous array of tables, indexes, and formulas manufactured by the testers gives the illusion of genuine ob-

jective reality. But these are only subjective assumptions dressed up in costumes. Our testing numerologists have as little relationship to real mathematics as did the astronomers of medieval times.

"Testing and Public Policy," American Psychologist, Vol. 20, November 1965, entire issue.

This issue presents a review of the controversy over testing as it emerged in Washington in 1965. The journal contains the testimony of witnesses at the Congressional hearings in June 1965. There were two such hearings: one for the Senate Subcommittee on Constitutional Rights of the Committee on the Judiciary, the other for the House Special Subcommittee on Invasion of Privacy. Some later comments from representatives of the American Psychological Association are included.

Testing, Testing, Testing. Report from Joint Committee on Testing. Washington, D.C., 1962.

This pamphlet contains a series of criticisms and recommendations made by the Joint Committee on Testing established by the American Association of School Administrators, the Council of Chief State School Officers, and the National Association of Secondary-School Principals.

Briefly, the following points are made.

1. The standardized test measures only a particular segment of performance most relevant to success in college. More attention should be given to other behavioral acts.

Pupils should be appraised on character and personality so that the individual can be helped to grow as a person. Furthermore, standardized ability tests fail to identify the

late bloomer and the creative student.

2. Objective ability measures also discriminate against those who are test-shy, emotionally disturbed, unmotivated, culturally deprived, or superior in ability.

3. Not all the education provided in secondary schools lends itself to objective assessment. In addition, a large number of individuals succeed in college despite their low ability scores.

4. There is much duplication in testing. A pupil in secondary school takes several aptitude tests in the eleventh and twelfth grades. Each of these, however, predicts the same criterion with the same validity.

5. What is tested in aptitude tests becomes so important that it ends up influencing school curriculums and teachers' judgments. What is not covered by ability tests gets left out of the curriculum. Thus large-scale testing programs may contribute, in certain respects, to the impairment of secondary education.

The committee recommends a set of procedures to provide a continuous and well-conceived plan of measurement and evaluation.

Trump, J. Lloyd, "What's Wrong with Testing," Twentieth Yearbook of the National Council on Measurement in Education. East Lansing: The Council, 1963, pp. 143-148.

Present difficulties with tests are more attributable to test users than to test constructors. Inadequate training, oversimplified explanations, and the economics of test publishing have all contributed to confusion. Background issues must be examined if the nature of the problem with testing is to be understood. These issues include devising better methods for evaluating excellence, specifying goals of in-

struction more precisely, avoidance of oversimplified criteria of achievement, development of richer descriptions of students, improved quality of local evaluation instruments, updating of tests to match changes in curriculums and instruction, provision of integrated systems of measurement, higher standards in test construction and administration.

Whyte, William H., Jr., The Organization Man. New York: Doubleday and Company, 1957, 471 pp.

In Whyte's terms, the organization man is the man who has taken over the spiritual values of organization life. He not only works for an organization; he belongs to it as well. The organization revolves about three major propositions: a belief in the group as a source of creativity; a belief in belongingness as the ultimate need of the individual; and a belief in the application of science to achieve the belongingness. This statement sets the general tone of the book.

Whyte devotes several chapters to the testing of the organization man, to show the organization's dependence on "these curious impositions into the psyche." These personality tests are not games; the individual must meet their demands in order to get ahead at the organization. As a result, the tests encourage conformity and submerge individual differences. They are not science, only the illusion of it.

It is a pathetic error to believe that tests can be scientific. They are enshrined in society's values. Though aptitude tests have proved useful in distinguishing capabilities, personality testing is fraught with imponderables. The tester tries to use statistics to convince people that he is translating uncertainty into certainty, the subjec-

tive into the objective. He has admirably succeeded in persuading organizations to use these personality tests for selecting employees and for granting promotion at high levels.

"If the layman gags at the phrasing of a question, testers reply, sometimes with a superior chuckle, this is merely a matter of 'face validity.' They concede that it is better if the questions seem to make sense, but they claim that the questions are not so important as the way large numbers of people have answered them over a period of time." What exactly does this mean? How do testers demonstrate the validity of the tests?

Test scores must be related to subsequent behavior of the people tested. But the problem is that when personality tests are used as selection devices, they become a large factor in the very equation they purport to measure. They screen out those who would upset the correlation, and therefore, for example, subsequent executives do not show certain personality profiles. The bias becomes institutionalized. The profile is self-confirming.

It is true that a first-grade organization requires a certain degree of homogeneity, but at the same time, these corporations must be prepared to respond to change. The sheer mechanics of testing punish the exceptional and far-seeing man. The intelligent mind sees shadings in the question, sets up alternatives, and finds it very difficult to answer a test with prefabricated choices.

If tests could, in fact, reveal the innermost self, would their use be justified? The author thinks not. In return for the salary, the organization can ask an individual for good work, but it should not ask for his psyche.

Womer, Frank B., "Pros and Cons of External Testing Programs," The North Central Association Quarterly, Vol. XXXVI, Fall 1961, pp. 201-210.

External testing programs are defined as those in which the results are used primarily by some institution or organization other than the school and in which the local school has no real choice as to whether its students take the tests. By this definition, the College Board tests, the American College Testing Program tests, and the National Merit Scholarship Corporation tests would be external examinations. What are the problems that have been associated with them?

One is that the pressures to participate are too great to be resisted and that these pressures have a detrimental effect. When the pressures on people and parents to obtain high scores become extreme, perhaps the whole relationship between the school and its constituents needs to be examined.

Another objection is that too many people are taking the tests, particularly in the case of the National Merit Scholarship Qualifying Test, a first scholarship screening test.

Too much school time is devoted to taking external tests. One would need to analyze internal as well as external testing to make a satisfactory judgment about this.

External testing is expensive.

The two major test publishers are more concerned with profits than with educational progress. The author says that it is true that competition exists, but this does not necessarily result in poor tests. It may, however, lead to duplication of testing.

It is claimed that external testing results in standardization of the school program, stifles experimentation, and

dictates teaching practices; but the fact is that we do not know how much effect external testing is having on high school education.

Proposed solutions have dealt with either a reduction in the number of testing programs or an alleviation of some of the problems associated with them. One suggestion is the establishment of equivalency tables for national tests. (The author favors this idea for guidance but not for selection or placement.) Another is aimed at improving communication with pupils and parents. Another calls for reducing publicity associated with becoming a scholarship semifinalist or winner, since this publicity leads to unfair comparisons between schools. Another is that the high schools take a strong stand against coaching for the external tests. A final suggestion is one made by some members of the Board staff. Board tests could be administered in June of the junior year for admissions and in June of the senior year for guidance and placement. This suggestion would seem to have considerable merit.